device, a dynaCT device, an angiogram device, and a mammography device, and the second modality is another of the MRI device, the CT device, the PET device, the ultrasound device, the dynaCT device, the angiogram device, and the mammography device.

[0020] In an embodiment, the first image and the second image are two-dimensional (2D) images, respectively.

[0021] In an embodiment, a method for machine training unsupervised multi-modal image registration in a medical imaging system includes defining a multi-task network with an objective function including a loss term representing latent space similarity and a loss term representing image space similarity. The multi-task network is machine trained to estimate a prediction of a deformation field for registration between a first image generated by a first modality and a second image generated by a second modality. The machine training is based on latent shape features in the latent space decomposed form multi-modal image pairs representing a region of interest. Images of each of the multi-modal image pairs have been generated by the first modality and the second modality, respectively.

[0022] In an embodiment, images of the multi-modal image pairs are not aligned.

[0023] In an embodiment, the method further includes storing, by a memory, the machine-trained multi-task network.

[0024] In an embodiment, the machine training is also based on translated images generated from the multi-modal image pairs being input into an image translation network.

[0025] In an embodiment, the method further includes registering the first image and the second image using the machine trained multi-task network.

BRIEF DESCRIPTION OF THE DRAWINGS

[0026] FIG. **1** shows a flowchart of one embodiment of a method for unsupervised multi-modal image registration;

[0027] FIG. **2** shows a flowchart of a method and corresponding network architecture of one embodiment for training an image translation network;

[0028] FIG. **3** illustrates examples of self-reconstruction loss for a multi-modal pair of images;

[0029] FIG. **4** shows a flowchart of a method and corresponding network architecture of one embodiment for training a deformable registration network;

[0030] FIG. **5** shows one embodiment of the generator of the deformable registration network;

[0031] FIG. **6** shows an example of a moving image, a fixed image, and a translated moving image based on a warping of one or more of the present embodiments; and

[0032] FIG. **7** shows one embodiment of a medical imaging system for image registration and/or therapy decision support.

DETAILED DESCRIPTION OF THE DRAWINGS

[0033] An unsupervised registration method for aligning intra-subject multi-modal images without ground truth deformation fields, aligned multi-modal image pairs, or any anatomical landmarks during training is provided. A parameterized registration function is learned via reducing a multi-modal registration problem to a mono-modal registration problem in latent embedding space.

[0034] For example, in one or more of the present embodiments, images are decomposed into a domain-invariant latent shape representation and a domain-specific appearance code based on the multi-modal unsupervised image-to-image translation framework (MUNIT). With the assumption that the intrinsic shape deformation between multi-modal image pairs is preserved in the domain-invariant shape space, an unsupervised diffeomorphic registration network is learned directly based on the disentangled shape representations. A similarity criterion may thus be defined in the latent space, minimizing a latent shape distance between a warped moving image and a target image.

[0035] A complimentary learning-based similarity metric is also provided. The complimentary learning-based similarity metric is defined via an adversarial loss to distinguish whether a pair of images is sufficiently aligned or not in the image domain. Since transformation is learned from a domain-invariant space, the method is directly applicable to bi-directional multi-modal registration without extra efforts (e.g., without landmark identification and/or supervised pairing).

[0036] Images of different modalities are embedded into a domain-invariant space via image disentangling, where any meaningful geometrical deformation may be directly derived in the latent space. The method includes three parts: an image disentangling network via unpaired image-to-image translation (e.g., an image translation network); a deformable registration network in the disentangled latent space; and an adversarial network.

[0037] FIG. **1** shows a flowchart of one embodiment of a method for unsupervised multi-modal image registration. The method may be performed using any number of imaging systems. The method is implemented in the order shown, but other orders may be used. For example, the method may not include use of the adversarial network.

[0038] The method is performed by medical imaging devices, a workstation, a server, a computer, or any combination thereof. The medical imaging devices or a memory of the medical imaging devices, the workstation, the server, and/or the computer are used to acquire data (e.g., image data) for a patient. An image processor, such as an image processor of the medical imaging devices, the workstation, the server, and/or the computer disentangle images and/or predict deformation fields. The image processor displays using a display screen or printer. A physician may use the output information to make a treatment decision for the patient.

[0039] In act **100**, a first image is acquired. The first image is generated by a first medical image modality. The first medical image modality may be any number of medical imaging devices including, for example, a magnetic resonance imaging (MRI) device, a computed tomography (CT) device, a positron emission tomography (PET) device, an ultrasound device, a dynaCT device, an angiogram device, a mammography device, or another type of medical imaging device.

[0040] The first image represents a region of interest within a patient. For example, the first image represents at least a portion of the liver of the patient. The first image may represent other regions of interest. In one embodiment, the first image is a three-dimensional (3D) image. In another embodiment, the first image is a two-dimensional (2D) image. For example, the first image may be a 2D image generated from a 3D volume. The first image may be of a particular size and/or resolution. For example, the first image is a 2D slice from a 3D volume with a size of